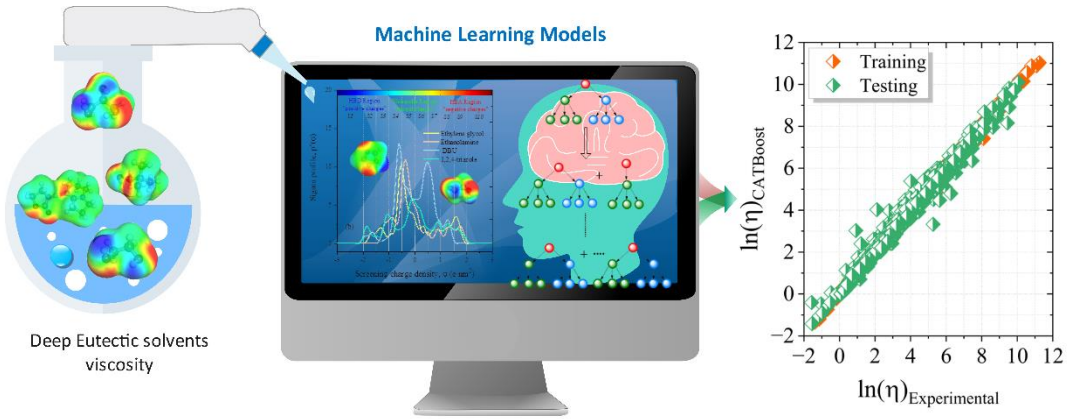
# Accurate Machine Learning for Predicting the Viscosities of Deep Eutectic Solvents




Schematic diagram for deep eutectic solvents viscosity predictions using the ML models based on the COSMO-RS-derived input features.

1. Mohan et al. (2024) "Accurate Machine Learning for Predicting the Viscosities of Deep Eutectic Solvents." J. Chem. Theory Comput., DOI: 10.1021/acs.jctc.3c01163

This work was supported and provided by the U. S. DOE Office of Science Genomic Science Program Office of Biological and Environmental Research (contract no. FWP ERKP752). Michelle K. Kidder acknowledges the U. S. DOE Office of Science, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences (CSGB) Grant Number 3ERKCG25 for partially supporting this research

## Scientific Achievement

We have shown that the charge polarization at the molecular surface obtained from a quantum chemical calculation (specifically, the COnductor-like Screening MOdel for Realistic Solvents [COSMO-RS]) can be used to train machine-learned (ML) regression models that accurately predict the Deep Eutectic Solvent (DES) viscosities.

## Significance and Impact

DESs are designer solvents that can be produced in unlimited combinations, it is experimentally infeasible to comprehensively measure the viscosities of all DESs of potential industrial interest. We demonstrate that we can achieve high accuracy from a featurization that does not require modeling any intermolecular complex geometries of HBA/HBD pairs, which is computationally expensive and likely requires generation of an ensemble of such bimolecular complex structures.

## Research Details

- In this work, a large dataset of 4949 data points was collected from the literature on the viscosity of 672 different DESs over wide ranges of temperature (from 278.15 K to 385.25 K) and molar ratios (HBA to HBD) from 1:19 to 19:1. Present study not only collected comprehensive dataset, but also covered the larger structural diversity of DESs with many clusters.
- Five ML models: multilinear regression (MLR), two-factor polynomial regression (two-factor PR), support vector regression (SVR), feed forward neural networks (FFNN), and categorical boosting (CatBoost) were developed to predict DES viscosity as a function of temperature and molar ratio.
- The COSMO-RS-derived sigma profile ($\sigma$-profile) features were used as inputs for the ML models.
- CatBoost showed the best predictive performance (lowest average absolute relative deviations and highest $R^2$ (0.987)) on the test set.